

**Supplementary information for:**

**Branched intermediate formation stimulates peptide bond  
cleavage in the GyrA intein protein splicing.**

Silvia Frutos, Michael Goger, Baldissera Giovani, David Cowburn and Tom  
W. Muir

## SUPPLEMENTARY METHODS

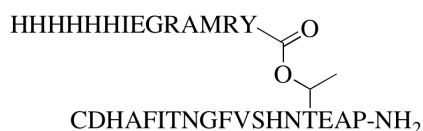
### General Methods

Nvoc-Thr-OH, Fmoc-Asn- $^{13}\text{CO}_2\text{H}$ , Fmoc-Phe- $^{13}\text{CO}_2\text{H}$ , Fmoc- $^{15}\text{NH}$ -Thr-OH and Fmoc- $^{15}\text{NH}$ -Val-OH were synthesized as described below. HBTU (2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate), HOBt (N-hydroxybenzotriazole) and all Fmoc amino acid derivatives were obtained from Novabiochem (San Diego, CA) except b-(2-thienyl)-alanine from Bachem (Torrance, CA) and unprotected labeled amino acids from Cambridge isotopes (Andover, MA). All other reagents were purchased from Sigma-Aldrich (St. Louis, MO). Analytical RP-HPLC was performed on a Hewlett-Packard 1100 series instrument equipped with  $\text{C}_{18}$  and  $\text{C}_4$  Vydac columns (5  $\mu\text{m}$ , 4.6 x 150 mm) at a flow rate of 1 mL/min using routine UV detection of 214 and 280 nm. Semi-preparative RP-HPLC was performed on a Waters DeltaPrep 4000 system with a Waters 486 tunable detector. All runs utilized a linear gradients of 0.1% TFA (trifluoroacetic acid) in water (solvent A) and 0.1% TFA 90% acetonitrile in water (solvent B). Mass spectrometric analysis for all peptides and proteins was performed by ESI-MS on a Sciex-API-100 single quadrupole spectrometer using positive ionization. Protein masses were calculated using the PeptideMass tool on the ExPASy server (<http://us.expasy.org>). Proteins were routinely analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) using standard protocols.

### Peptide Synthesis

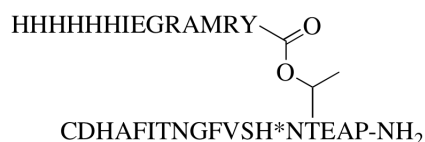
All peptides were synthesized manually using standard Fmoc SPPS (solid phase peptide synthesis) on Rink-amide ChemMatrix resin (0.5 mmol/g, Matrix Innovations, Canada). Linear chain assembly was carried out with HBTU activation (4.8 eq) using a 5-fold excess of amino acid over the resin in DMF (dimethylformamida) with DIEA (N,N-diisopropylethylamine). Branched chain assembly was performed with HOBt and DICP (N,N'-diisopropylcarbodiimide) activation using 5 fold-excess of amino acid to resin in DMF. A stream of dry  $\text{N}_2$  was used to agitate the reaction mixture. Double coupling was routinely used to ensure quantitative acylations after b-branched amino acids. The Fmoc protecting group was removed with 20% piperidine in DMF (1 x 2 minutes, followed by 2 x 10 minutes) for the linear chain or 2% DBU (1,8-diazabicyclo [5.4.0] undec-7-ene)

for the branched chain. Peptidyl-resin was washed between coupling cycles with DMF for 3 minutes by alternating batch or flow washes. Peptides were cleaved from the resin using 94% TFA, 1% triisopropylsilane (TIS), 2.5% ethanedithiol, and 2.5% H<sub>2</sub>O (cleavage cocktail) (**Supplementary Fig. 3**). Crude peptide products were precipitated and washed with cold Et<sub>2</sub>O, dissolved in solvent A with a minimal amount of solvent B and then purified by RP-HPLC and characterized by ESI-MS (**Supplementary Fig. 4**).



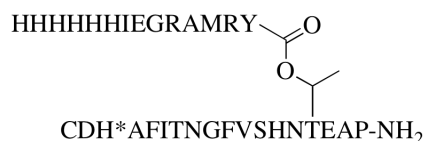
RP-HPLC (C<sub>18</sub> column, gradient of 20-40% solvent B) and the molecular weight confirmed by ESI-MS (observed 2782.18 ± 1.61, expected 2782.95 Da). The isolated yield after synthesis and purification was ~17%.

### Synthesis of peptide used in the semi-synthesis of construct **5**



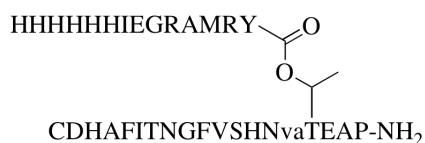
Branched peptide assembly was carried out as described above for the wt peptide **1**, except Fmoc-His<sup>197</sup>(Trt)-OH was replaced by Fmoc-β-(2-thienyl)-Ala-OH (abbreviated as H\*). The peptide was purified by RP-HPLC (C<sub>18</sub> column, gradient of 20-40% solvent B) and the molecular weight confirmed by ESI-MS (observed 3775.7 ±0.73, expected 3774.94 Da). The isolated yield after synthesis and purification was ~8%.

### Synthesis of peptide used in the semi-synthesis of construct 6



Branched peptide assembly was carried out as described above for the wt peptide **1**, except Fmoc-His<sup>187</sup>(Trt)-OH was replaced Fmoc-β-(2-thienyl)-Ala-OH (abbreviated as H\*). Following cleavage, the crude peptide was purified by RP-HPLC (C<sub>18</sub> column, gradient of 20-40% solvent B) and the molecular weight confirmed by ESI-MS (observed 3774.82 ± 1.53, expected 3774.94 Da). The isolated yield after synthesis and purification was ~15%.

### Synthesis of peptide used in the semi-synthesis of construct 7

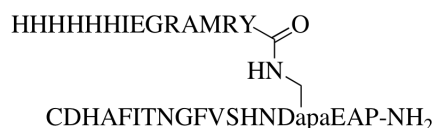


Branched peptide assembly was carried out as described above for the wt peptide **1**,



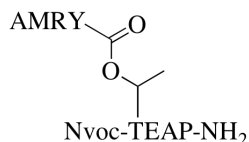
except Fmoc-Asn(Trt)-OH was replaced by Fmoc-Nva-OH at Asn198. Following cleavage, the crude peptide was purified by RP-HPLC ( $C_{18}$  column, gradient of 20-40% solvent B) and the molecular weight confirmed by ESI-MS (observed  $3744.29 \pm 0.5$ , expected 3744.64 Da). The isolated yield after synthesis and purification was ~13%.

#### Synthesis of peptide used in the semi-synthesis of construct 8



Linear chain assembly was carried out as described above for the wt peptide **1**, except that Fmoc-Thr-OH was replaced by Fmoc-Dapa(Mtt)-OH at Thr+1 and Boc-Thz-OH was used in place of Boc-Cys(Trt)-OH at Cys184. Following linear chain elongation, the Mtt protecting group was removed by treatment with 1% TFA, 5% TIS in DCM (5 x 2 min). Fmoc-Tyr(Trt)-OH was coupled using HBTU and DIEA. The remaining residues were coupled using HBTU and DIEA as above with 20% piperidine in DMF (2 x 10 min) used to remove the Fmoc protecting groups. Following cleavage, the lyophilized crude peptide was added to 0.5 M methoxylamine in water and the solution agitated for 6 hours at 37 °C. The desired product was the major product (~95%) as determined by RP-HPLC. The peptide was purified by RP-HPLC ( $C_{18}$  column, gradient of 20-40% solvent B) and the molecular weight confirmed by ESI-MS (observed  $3744.48 \pm 0.9$ , expected 3744.32 Da). The isolated yield after synthesis and purification was ~14%.

#### Synthesis of peptide 9



Linear chain Nvoc-Thr<sup>+1</sup>-Pro<sup>+4</sup> (residue numbers refer to GyrA extein residues) was assembled using the synthetic methods described above, except that Nvoc-Thr-OH was

used in place of Fmoc-Thr-OH. Undesired Thr side chain acylation was cleaved with ethanolamine in 5% water/DMF (2 x 30 min). The branch sequence was then added as peptide **1**. Following cleavage, the crude peptide was purified by RP-HPLC (C<sub>18</sub> column, gradient of 15-40% solvent B) and the molecular weight confirmed by ESI-MS (observed 1176.9, expected 1175.49 Da). The isolated yield after synthesis and purification was ~19%.

#### Synthesis of Nvoc-Thr-OH

L-Thr (200 mg, 0.279 mmol) was dissolved in a mixture of water (50 mL) and Et<sub>2</sub>O (50 mL). NaHCO<sub>3</sub> (610 mg, 0.725 mmol) and 4,5-dimethoxy-2-nitrobenzyl chloroformate (500 mg, 0.181 mmol) were added at 0 °C and the mixture stirred at room temperature for 2 h. Additional NaHCO<sub>3</sub> (610 mg, 0.725 mmol) and 4,5-dimethoxy-2-nitrobenzyl chloroformate (500 mg, 0.181 mmol) were added and the mixture stirred at room temperature for an additional 18 h. After removal of the organic layer by evaporation under reduced pressure, the residual aqueous layer was acidified to pH 2 and extracted with ethyl acetate (30 mL). The organic phase was dried over anhydrous Na<sub>2</sub>SO<sub>4</sub> and the solvent evaporated under reduced pressure. The crude compound was applied to preparative C<sub>18</sub> RP-HPLC and the products eluted with a linear gradient of CH<sub>3</sub>CN in 0.1% aqueous TFA. The desired fraction was collected and lyophilized (yield 78.2%). NMR (400 MHz, DMSO-*d*<sub>6</sub>) <sup>1</sup>H: δ 7.71 (s, 1H), 7.38 (d, 1H), 7.26 (s, 1H), 5.41 (d, 1H), 5.31 (d, 1H), 4.14-3.96 (m, 1H), 3.97 (s, 3H), 3.91 (s, 3H), 1.12 (d, 3H); <sup>13</sup>C NMR: δ 173.12, 156.98, 154.46, 148.47, 139.76, 129.17, 110.82, 63.42, 60.80, 57.00, 56.71, 21.34, 21.22; ESI-MS expected (M+H<sup>+</sup>) 359.15, observed 359.0.

#### Synthesis of H<sub>2</sub>N-Asn-<sup>13</sup>CO<sub>2</sub>H

H<sub>2</sub>SO<sub>4</sub> (0.25 mL) was added to a suspension of L-Asp [<sup>13</sup>CO<sub>2</sub>H] (500 mg) in dry MeOH (5 mL) and stirred for 4 h at 40-45 °C. The solution was cooled to 25 °C and treated with 28% aqueous ammonia to a pH of 3.5. The mixture was concentrated under vacuum to give a thick slurry then treated with 11.3 mL of 28 % aqueous ammonia and 0.3 mL H<sub>2</sub>O and stirred at room temperature overnight. The mixture was concentrated under vacuum until nearly dry, redissolved in 0.5 mL H<sub>2</sub>O at 55 °C, adjusted to pH 5.4, and then cooled to 25 °C. The solid was collected by filtration, washed and dried, and finally

recrystallized to yield 0.34 g (yield 60%).  $^1\text{H}$  NMR (400 MHz,  $\text{D}_2\text{O}$ )  $\delta$  3.96-3.90 (m, 1H), 3.03-2.759 (m, 2H);  $^{13}\text{C}$ :  $\delta$  174.80, 173.61, 144.17, 51.63, 34.83. ESI-MS expected ( $\text{M}+\text{H}^+$ ) 135.12, observed 135.0.

#### Synthesis of Fmoc-labeled amino acids

$\text{NaHCO}_3$  (2.00 mmol) and the appropriate isotopically labeled amino acid (1.26 mmol) were dissolved in a minimal amount of water ( $\sim 5$  mL). *1,4*-dioxane (10 mL) was added followed by a solution of Fmoc-OSu (1.25 mmol) in *1,4*-dioxane over 15 min. The reaction was stirred at r.t. for 20 h. At this time, the reaction was acidified to pH 3 using HCl and extracted three times with EtOAc. Organic layers were combined, dried over  $\text{Na}_2\text{SO}_4$ , filtered, and concentrated. The crude compound was applied to preparative  $\text{C}_{18}$  RP-HPLC and eluted with a linear gradient of  $\text{CH}_3\text{CN}$  in 0.1% aqueous TFA. The desired fraction was collected and lyophilized (**Fmoc- $^{15}\text{NH}$ -Thr-OH**  $\sim 85\%$ , purity 98%. NMR (400 MHz,  $\text{DMSO}-d_6$ )  $^1\text{H}$ :  $\delta$  7.89 (d, 2H), 7.75 (d, 2H), 7.43 (t, 2H), 7.32 (t, 2H), 7.07 (d, 1H), 4.32-4.22 (m, 3H), 4.10 (m, 1H), 3.97-2.94 (m, 1H), 1.11 (d, 3H);  $^{13}\text{C}$  NMR:  $\delta$  173.15, 157.27, 144.71, 141.59, 128.52, 127.94, 126.24, 120.92, 67.30, 60.80, 47.51, 35.28, 21.26; ESI-MS expected ( $\text{M}+\text{H}^+$ ) 343.4, observed 343.0. **Fmoc-Phe- $^{13}\text{CO}_2\text{H}$**   $\sim 90\%$ , NMR (400 MHz,  $\text{DMSO}-d_6$ )  $^1\text{H}$ :  $\delta$  7.87 (d, 2H), 7.71 (d, 1H), 7.64 (t, 2H), 7.40 (t, 2H), 7.33-7.19 (m, 8H), 4.20-4.16- (m, 4H), 3.10-3.06 (m, 1H), 2.90-2.84 (m, 1H);  $^{13}\text{C}$ :  $\delta$  174.17, 156.79, 144.61, 141.54, 138.86, 129.96, 129.03, 128.48, 127.91, 127.22, 126.19, 66.47, 56.31, 37.33; ESI-MS expected ( $\text{M}+\text{H}^+$ ) 389.4, observed 389.0. **Fmoc-Asn- $^{13}\text{CO}_2\text{H}$**   $\sim 74\%$ , NMR (400 MHz,  $\text{DMSO}-d_6$ )  $^1\text{H}$ :  $\delta$  7.89 (d, 2H), 7.71 (d, 2H), 7.51 (d, 1H), 7.43 (t, 2H), 7.33 (t, 2H), 4.52-4.22 (m, 4H) 2.59-2.32 (m, 2H).  $^{13}\text{C}$ :  $\delta$  174.03, 172.05, 156.66, 144.67, 141.57, 128.51, 127.97, 126.21, 121.03, 66.56, 51.42, 47.47, 37.63; ESI-MS expected ( $\text{M}+\text{H}^+$ ) 357.4, observed 357.0. **Fmoc- $^5\text{NH}$ -Val-OH**  $\sim 84\%$ , NMR (400 MHz,  $\text{DMSO}-d_6$ )  $^1\text{H}$ :  $\delta$  7.89 (d, 2H), 7.75 (d, 2H), 7.59 (d, 1H), 7.43 (t, 2H), 7.32 (t, 2H), 4.28-4.20- (m, 2H), 3.90-3.86 (m, 1H), 2.09-2.04 (m, 1H), 0.92 (d, 6H);  $^{13}\text{C}$ :  $\delta$  174.13, 157.28, 144.73, 141.57, 128.51, 127.92, 126.23, 121.01, 66.61, 60.47, 47.53, 30.46, 20.08, 19.04. ESI-MS expected ( $\text{M}+\text{H}^+$ ) 341.4, observed 341.0.

#### **Construction of Plasmids**

The DNA encoding the *Mxe* GyrA intein (residues 1-184) was PCR amplified from the

pTXB1 vector to incorporate an additional N-terminal His<sub>6</sub>-tag and the Factor Xa recognition site. This PCR product was re-cloned back into the NdeI and SapI restrictions sites of the pTXB1 vector, which contain a mutated *Mxe* GyrA intein and the chitin-binding domain. Note, mutation of Cys1 was also found to be required for efficient production of the recombinant intein  $\alpha$ -thioester fragments required for the EPL reaction. The C1S mutation was then introduced using Quick-Change mutagenesis kit (Stratagene) to generate the final fusion protein (His<sub>6</sub>-Xa-Ser1-(2-184)-GyrACBD) used to prepare branched constructs. To prepare the fusion protein (His<sub>6</sub>-Xa-AMRY-Ala1-(2-184)-GyrACBD) used in the linear construct **3**, Quick-Change mutagenesis kit was used to introduce four N-extein residues (AMRY) and to mutate the Ser1 to Ala.

### **Protein Expression**

*E. coli* BL21(DE3) cells (Novagen, Darmstadt, Germany) were transformed with the appropriate plasmid and grown at 37 °C either in M9 minimal media supplemented with [<sup>15</sup>N]NH<sub>4</sub>Cl for NMR labeling experiments or LB (*Luria Bertani*) for kinetic studies. Upon reaching an OD<sub>600</sub> ~0.8, protein expression was induced with 0.5 mM IPTG (Isopropyl  $\beta$ -D-1-thiogalactopyranoside) at 37 °C for 3 h. The cells were centrifuged and the pellet suspended in five volumes of lysis buffer (50 mM Tris-HCl pH 7.9, 2 mM EDTA, 500 mM NaCl) supplemented with one complete protease inhibitor tablet EDTA free (Roche, Basel, Switzerland) per 50 mL of buffer. Cells were lysed by passage through a French press (Emulsiflex C5 homogenizer, Avestin, Ottawa, Canada) and the inclusion bodies isolated by centrifugation at 16,000 rpm for 30 minutes in an SS-34 rotor. The insoluble protein was re-suspended and washed twice with lysis buffer containing 0.1 % Triton X-100 and washed once with lysis buffer alone to remove trace amounts of Triton X-100. The resulting pellet was dissolved in 100 mL of re-suspension buffer (50 mM Tris-HCl pH 7.9, 6 M urea and 500 mM NaCl) and loaded onto a Ni-NTA affinity column. The resin was washed with re-suspension buffer and the fusion protein eluted from the column in 10 mL fractions with re-suspension buffer supplemented with 500 mM imidazole. The fractions were analyzed by SDS-PAGE and those containing the 50 kDa fusion protein were pooled.

### **Generation of protein $\alpha$ -thioester**

Fusion proteins were refolded by step-wise dialysis from a buffer containing, 50 mM Tris-HCl pH 7.6, 100 mM NaCl, 1 mM EDTA and 6 M urea against 50 mM Tris-HCl pH 7.6, 100 mM NaCl and 1 mM EDTA at 4 °C. Thiolysis was initiated by adding sodium 2-mercaptoethane sulfonate (MESNA) to a final concentration of 200 mM and allowed to proceed for 2 days at 25 °C. The thioester-containing cleavage products were purified by preparative scale RP-HPLC on a C<sub>4</sub> column using a gradient of 40-65% buffer B and fractions analyzed by analytical RP-HPLC and ESI-MS. Pure thioester products were pooled, lyophilized and then refolded by step-wise dialysis from a buffer containing, 50 mM Tris-HCl pH 7.6, 100 mM NaCl and 6 M urea against 50 mM Tris-HCl pH 7.6, 100 mM NaCl and 100 mM MESNA at 4 °C. The N-terminal His<sub>6</sub>-tag was proteolytically removed with Factor Xa (1U enzyme/500 mg protein for 6 h at 25 °C). The final thioester products were purified by preparative scale RP-HPLC on C<sub>4</sub> column using a gradient of 40-65% buffer B and analyzed by analytical RP-HPLC and ESI-MS. ~30 mg of pure thioester-containing products were obtained from 1L of *E. coli* culture (**Supplementary Fig. 2**).

### **Expressed protein ligation**

Expressed protein ligation (EPL) was performed at 4 °C by mixing the appropriate synthetic peptide (3 eq.) and protein  $\alpha$ -thioester (1 eq, 1 mM) in ligation buffer (100 mM NaPi pH 7.8, 6M Gmd.Cl 100 mM NaCl) plus 100 mM MESNA and 10 mM TCEP (tris[2-carboxyethyl] phosphine). The reaction was followed by analytical RP-HPLC and ESI-MS. All reactions were >90% complete within 5 days. Semi-synthetic proteins were separated from unreacted precursors on a Ni-NTA affinity column. The resin was washed with buffer (100 mM Tris-HCl pH 7.5, 6M Urea, 100 mM NaCl and 1 mM TCEP), the protein eluted from the column in 8 x 0.5 mL fractions of buffer supplemented with 500 mM imidazole and the fractions analyzed by SDS-PAGE. Finally, the ligated products were purified by semi-preparative scale RP-HPLC on a C<sub>4</sub> column using a gradient of 40-50% solvent B and analyzed by analytical RP-HPLC and ESI-MS (yield 30%, 2 mg). (**Supplemental Fig. 5 & 6**).

### **Nuclear magnetic resonance (NMR) spectroscopy**

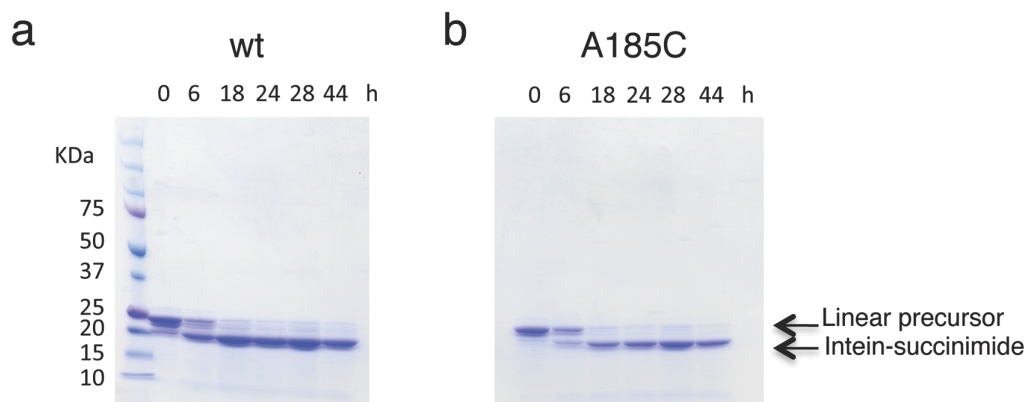
Purified constructs were folded at 4 °C by stepwise dialysis from denaturing buffer into

the buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl and 1 mM TCEP). Two-dimensional  $^1\text{H}$ - $^{15}\text{N}$  HSQC (heteronuclear single-quantum correlation spectra) of isotopically labeled samples were collected on a Bruker Avance-800 or 900 spectrometers at 4°C. For the HSQC experiments, we collected 256 complex points in the  $^1\text{H}$  and 128 in the  $^{15}\text{N}$  dimensions. In the 1D HNCO experiments we collected 1024 complex points in the  $^1\text{H}$  dimension and 1 complex point in the  $^{13}\text{C}$  and  $^{15}\text{N}$  dimension. The chemical shifts for the Phe194-Val195 peptide bond were  $^1\text{H}$   $\delta$ =7.89 ppm,  $^{15}\text{N}$   $\delta$ =116.69 ppm, and  $^{13}\text{C}$   $\delta$ =171.71 ppm at pH 7.5 and  $^1\text{H}$   $\delta$ =7.89 ppm,  $^{15}\text{N}$   $\delta$ =116.44 ppm, and  $^{13}\text{C}$   $\delta$ =172.792 ppm at pH 4.5. The chemical shifts for the Asn198-Thr+1 peptide bond were  $^1\text{H}$   $\delta$ =9.22 ppm,  $^{15}\text{N}$   $\delta$ =121.60 ppm, and  $^{13}\text{C}$   $\delta$ =171.98 ppm at pH 7.5 and  $^1\text{H}$   $\delta$ =8.29 ppm,  $^{15}\text{N}$   $\delta$ =113.52 ppm, and  $^{13}\text{C}$   $\delta$ =172.53 ppm at pH 4.5 (**Supplementary Fig. 12**). Data sets were multiplied by a cosine-bell window function and zero-filled to 1,000 points before Fourier transformation using XWINNMR (Bruker Instruments). The corresponding sweep widths were 11.9 ppm, 12.8 ppm, and 33.3 ppm in the  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  dimensions, respectively. Experimental amide  $^1J_{\text{NC}}$  coupling constants were obtained by fitting the time evolution of the normalized peak intensities extracted from the series of HNCO-type experiments using the expression

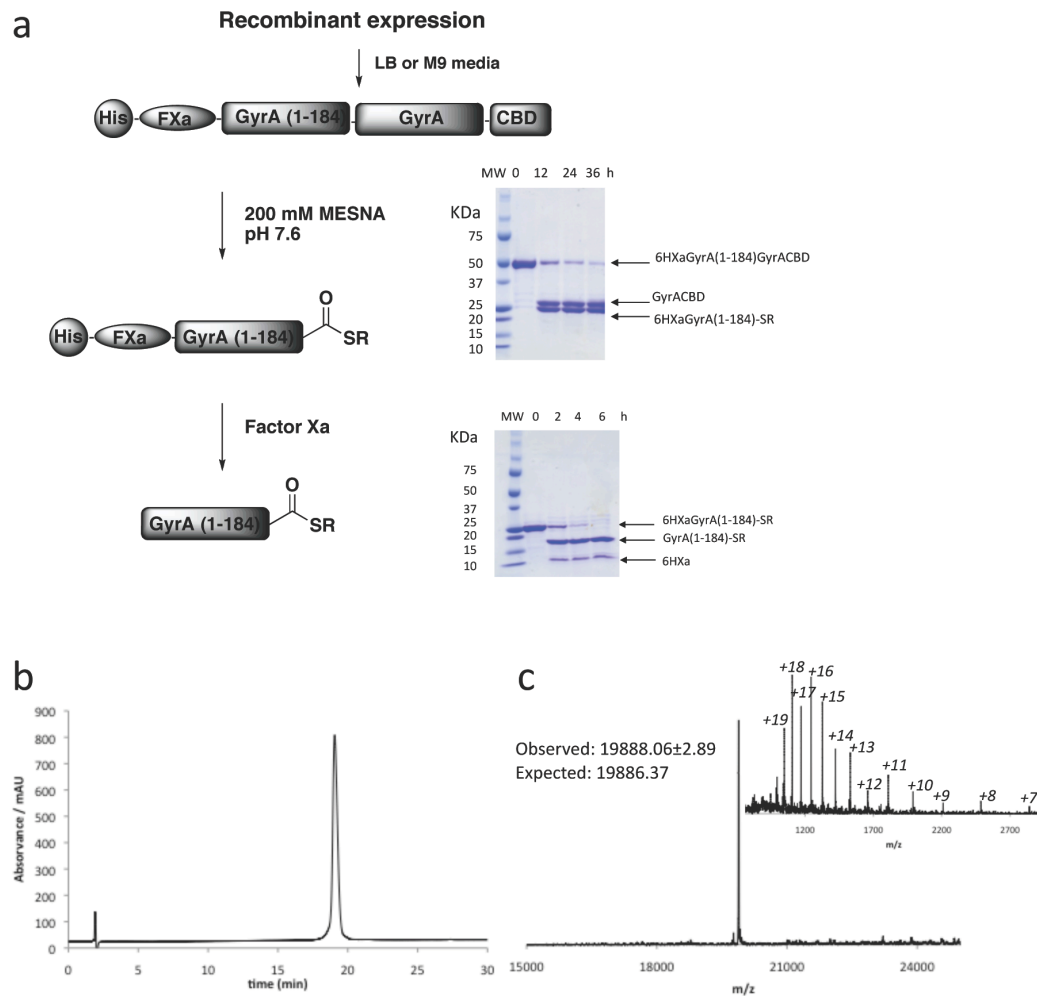
$$I_k = e^{-4t_1 R_{2k}} \sin^2(2\pi J_k t_1)$$

where  $I_k$  are the normalized peak intensities for peak  $k$ ,  $R_2$  is the transverse relaxation time for peak  $k$ , and  $t_1$  is the indirect dimension delay.

## SUPPLEMENTARY RESULTS

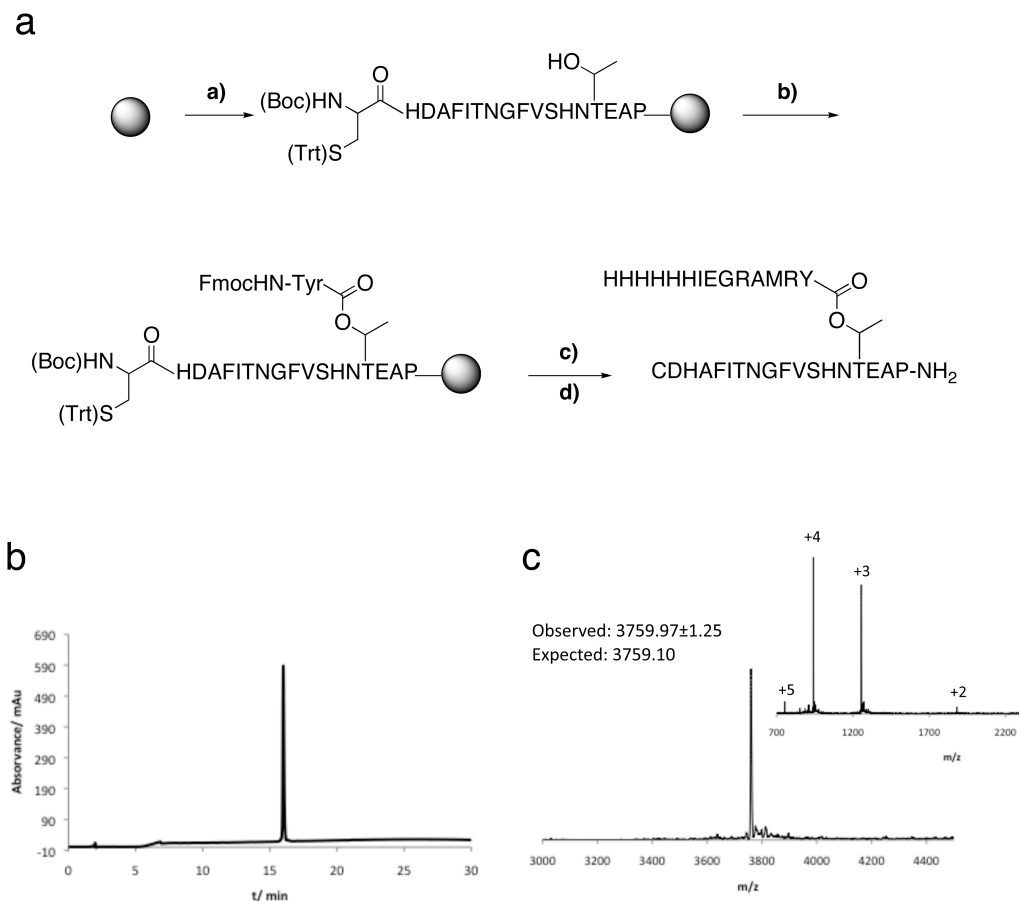


**Fig. S1:** Coomassie stained SDS-PAGE analysis of splicing reactions for the wild-type (a) and mutant A185C (b) *Mxe* GyrA inteins each embedded in a native extein context. Bands corresponding to the linear splicing precursor and intein-succinimide product are indicated.

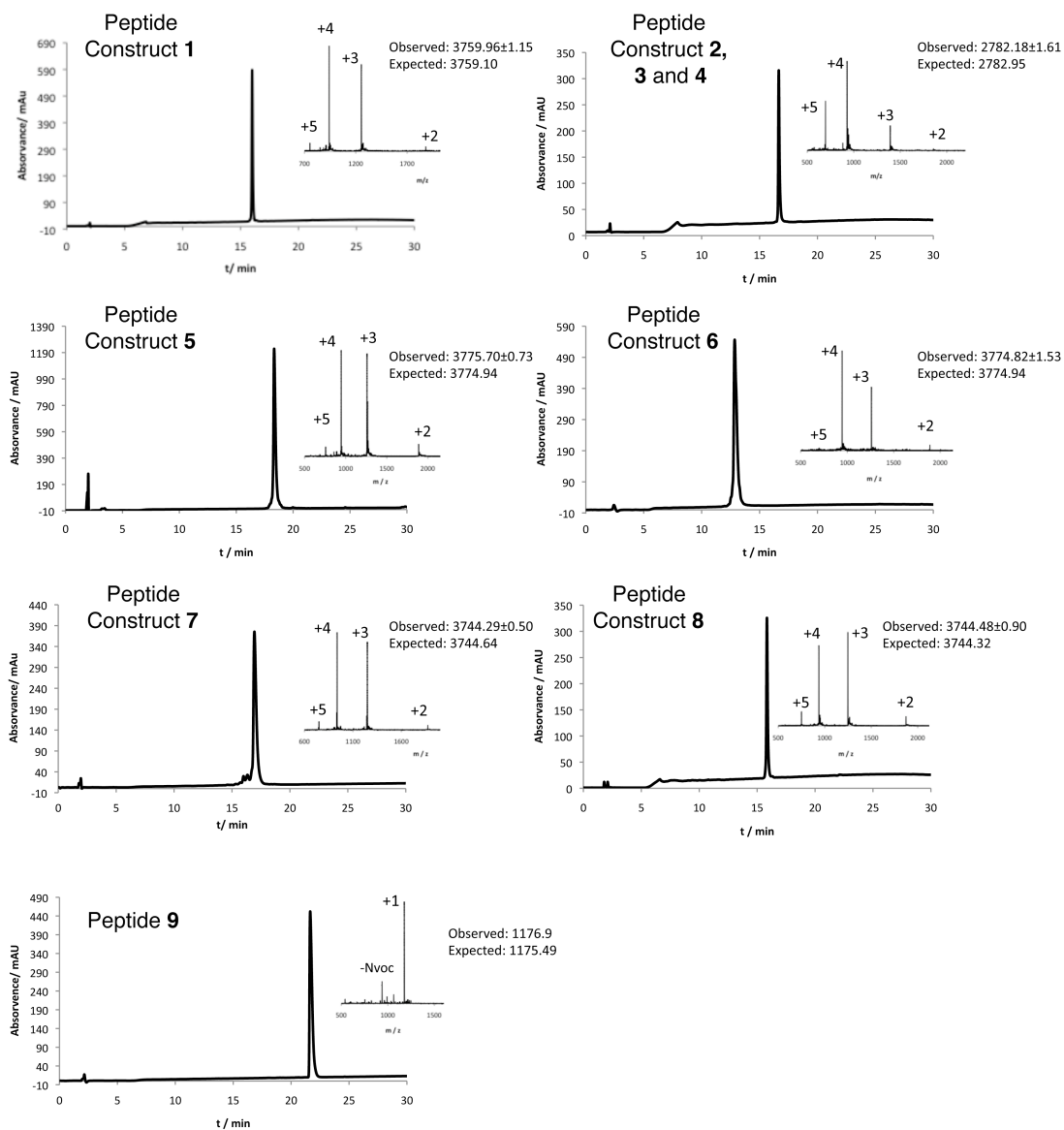


**Fig. S2:** Generation of *Mxe* GyrA intein thioester fragment. **(a)** Schematic of the protocol used to generate the recombinant protein  $\alpha$ -thioester. At right are coomassie stained SDS-PAGE gels of the thiolysis and proteolysis reactions. **(b)** RP-HPLC analysis of purified *Mxe* GyrA intein (1-184)  $\alpha$ -thioester ( $C_{18}$  column, 35-60% B over 30 min, 214 nm detection). **(c)** ESI-MS analysis of the purified thioester.

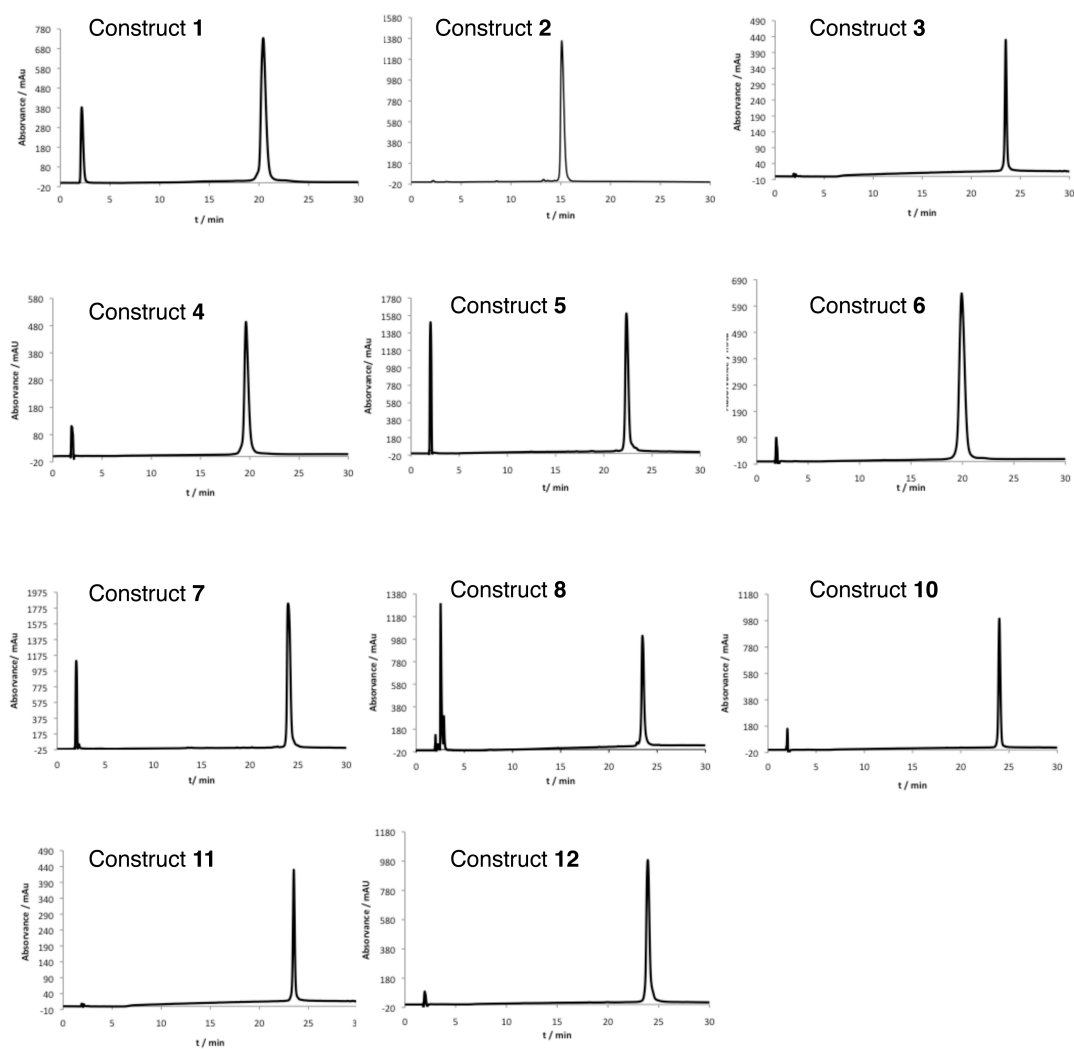




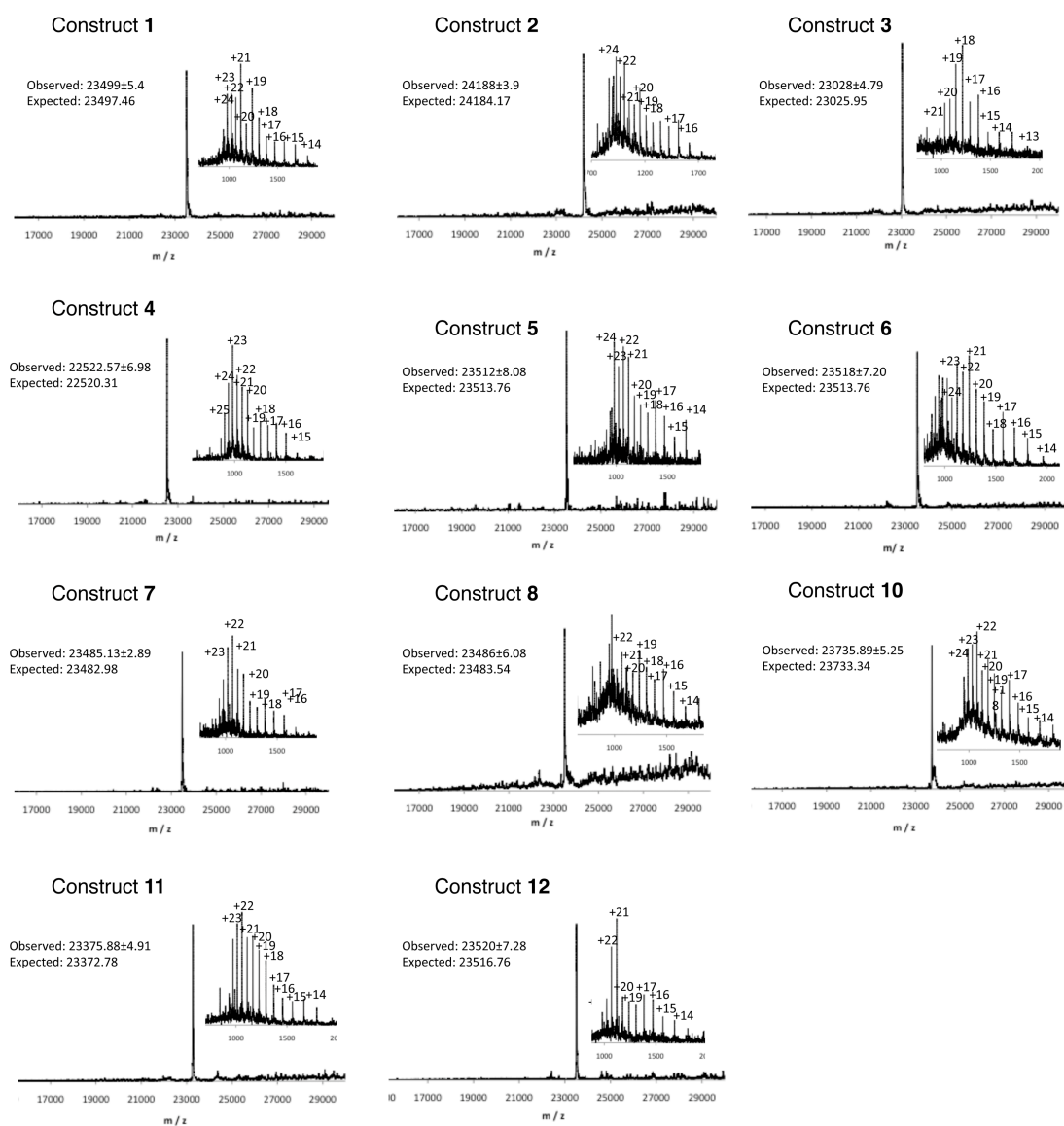
**Fig. S3:** Solid phase synthesis of branch peptide. **(a)** Schematic of synthetic route. Reaction conditions: a) Standard Fmoc-SPPS. b) Ethanolamine in 5% water/DMF (2 x 30 min) followed by ester formation using 4 eq. of Fmoc-Tyr(<sup>t</sup>Bu)-OH, 4 eq. (1-(2-mesitylenesulfonyl)-3-nitro-1H-1,2,4-triazole) and 12 eq. of N-methylimidazole(1 x O/N). c) Fmoc-SPPS using 5 eq. of amino acid, 5 eq. of HOBT and 5 eq. of DIC and 2% of DBU in DMF to remove the Fmoc. d) 1% triisopropyl silane, 2.5% ethanedithiol, 2.5% water and 94% trifluoroacetic acid, 1.5 h, r.t. **(b)** RP-HPLC analysis of purified branched peptide (C<sub>18</sub> column, 30-70% B over 30 min, 214 nm detection). **(c)** ESI-MS analysis of the branched peptide .



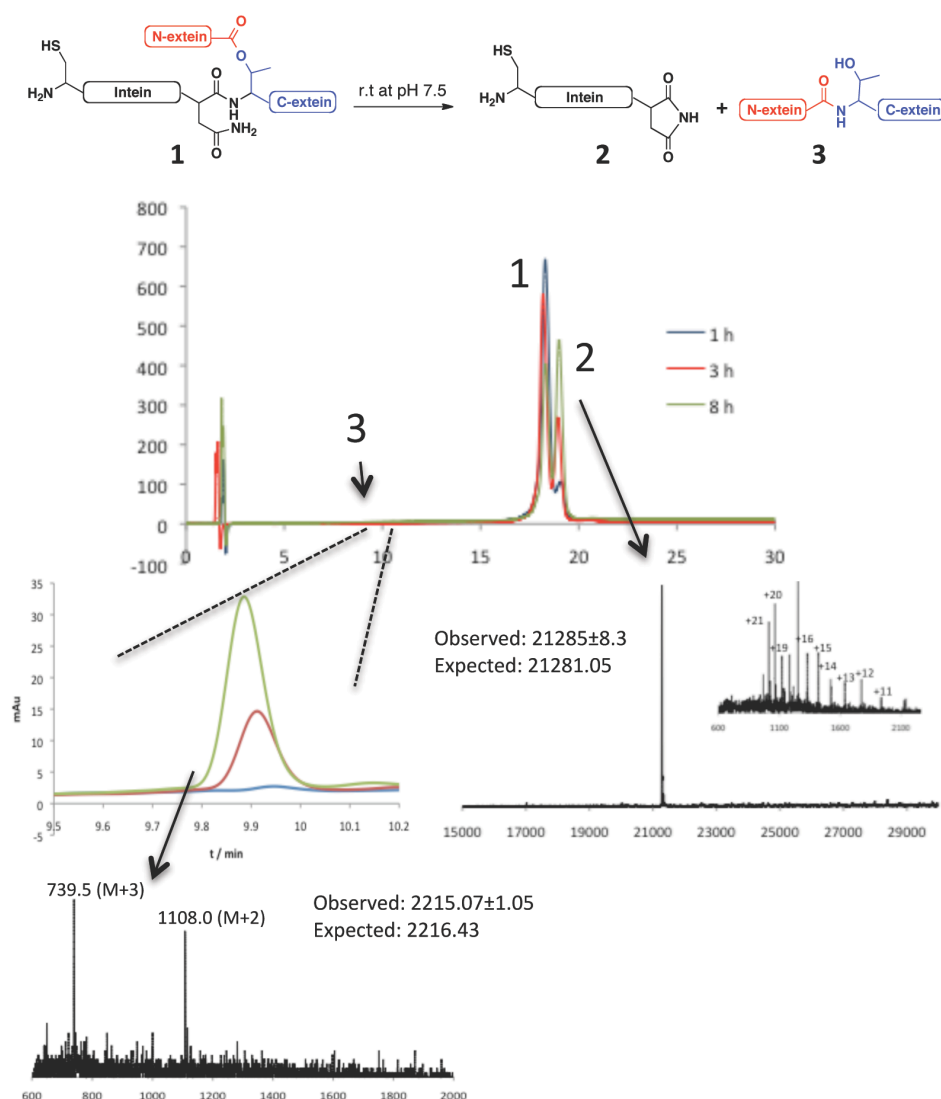
**Fig. S4:** RP-HPLC and ESI-MS analysis of purified synthetic peptide 1-9 ( $C_{18}$  column, 214 nm detection).



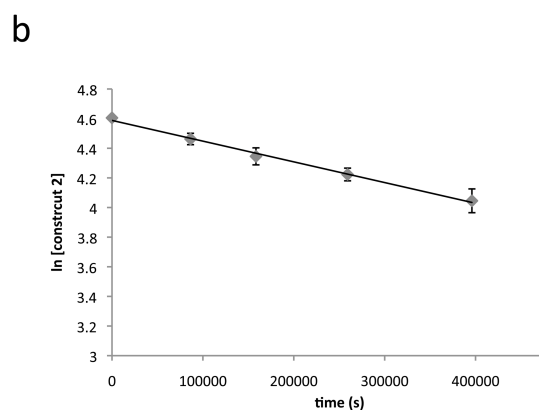
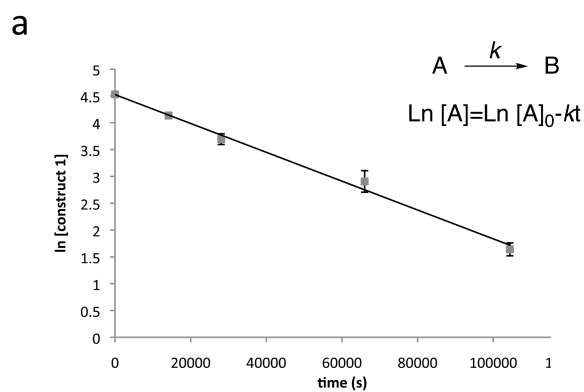
**Fig. S5:** RP-HPLC analysis of purified semi-synthetic proteins **1-8 & 10-12** ( $C_{18}$  column, 35-60 % B over 30 min or 30-60% B over 30 min., 214 nm detection).



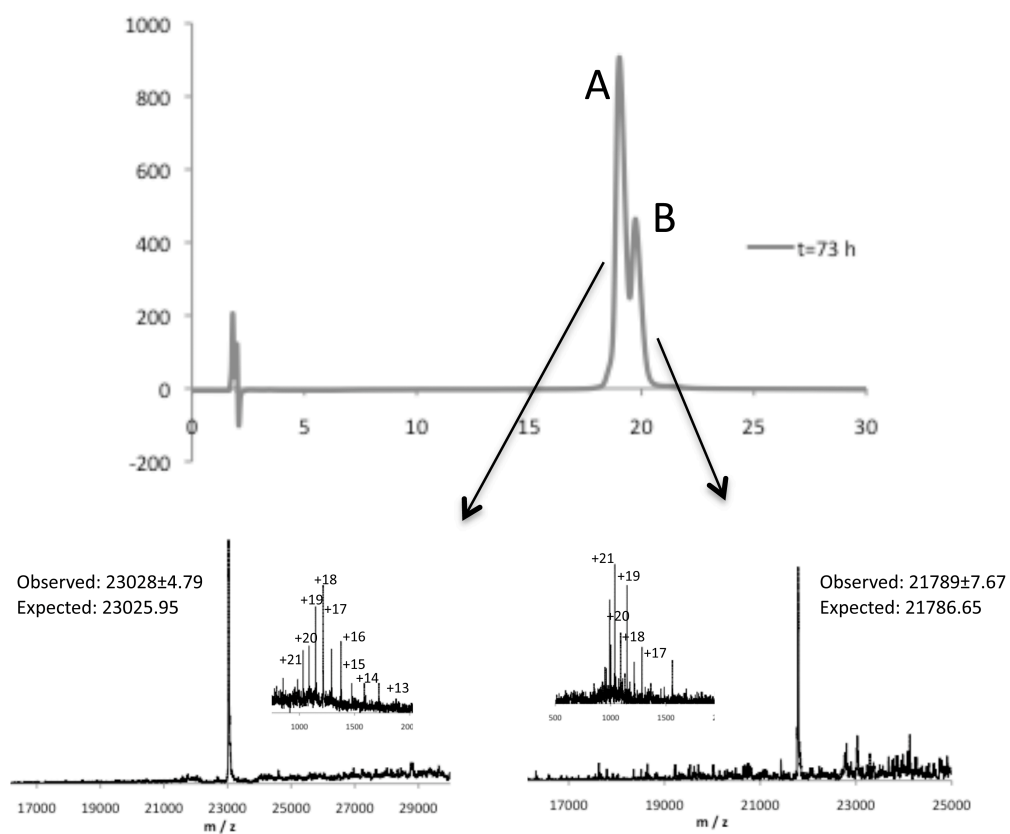
**Fig. S6: ESI-MS analysis of the semi-synthetic constructs 1-8 & 10-12.**



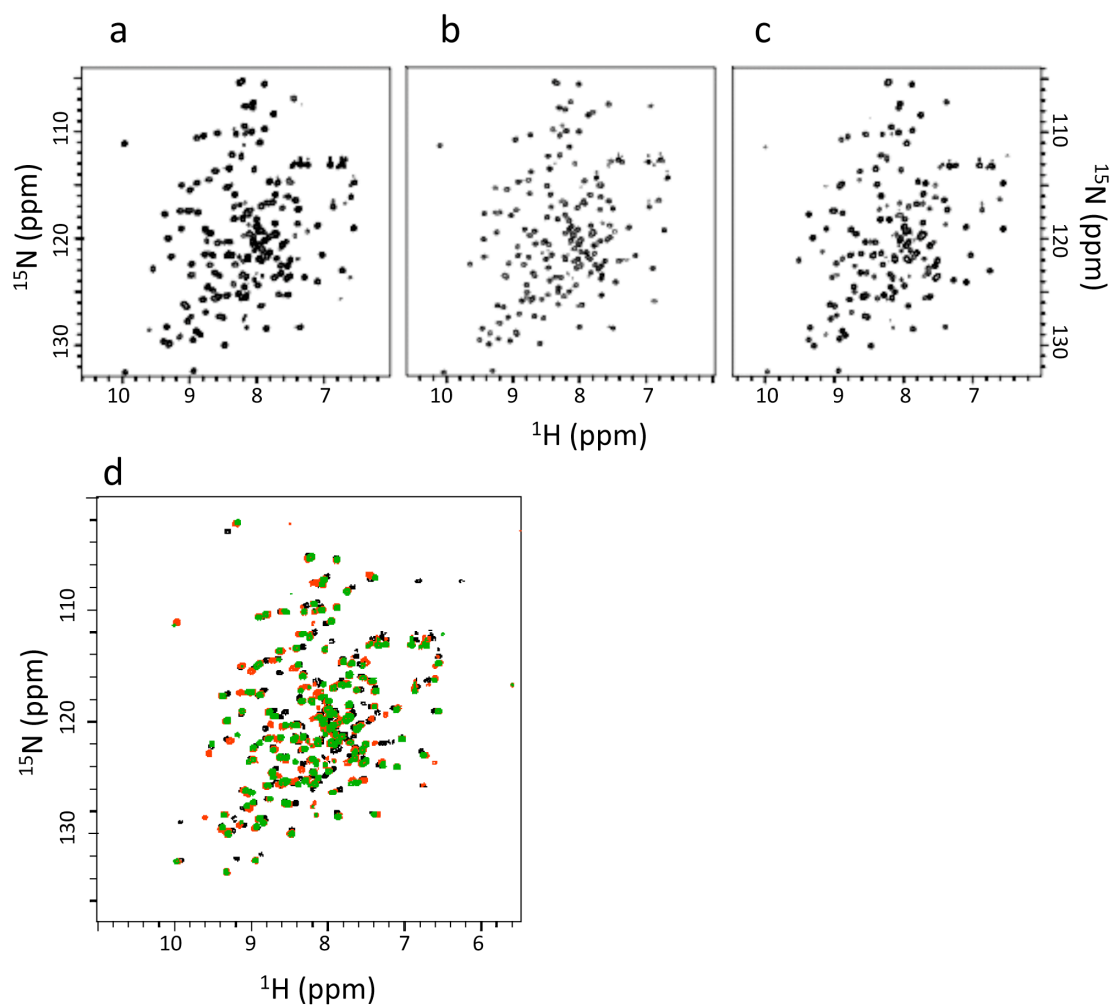
**Fig. S7:** Analysis of splicing reactions of construct **1**. Succinimide product (**2**) and spliced product (**3**) formation of the branched construct **1** was monitored by HPLC over time at 25 °C and pH 7.5. The composition of the different peaks was confirmed by mass spectrometry.



**Fig. S8:** Kinetic analysis of splicing reactions. Succinimide formation was monitored by plotting the  $\ln$  of the concentration of the branched construct **1** (**a**) and the linear analog **3** (**b**) over time at 25 °C and pH 7.5. The data from the time course were fitted to a first-order rate function to obtain the observed rate constants shown in Table 1.

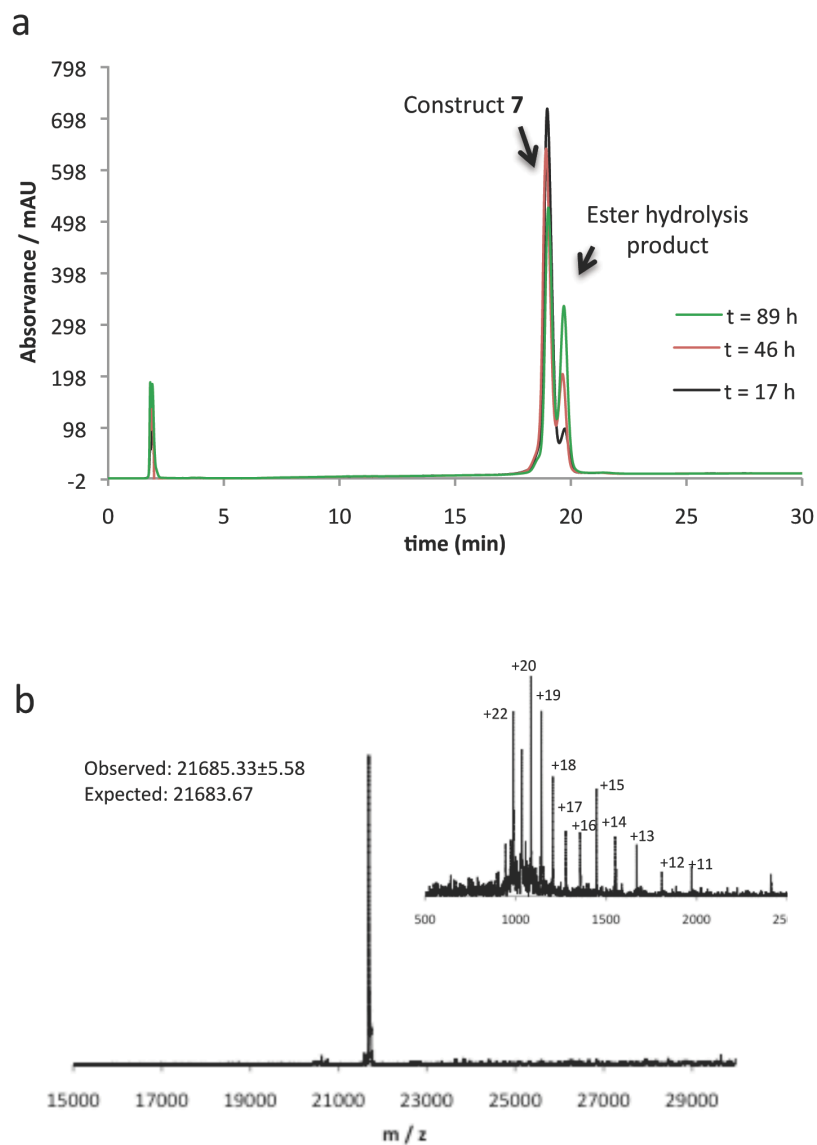


**Fig. S9:** Analysis of splicing reactions of construct **3**. Succinimide product (B) formation of the lineal construct **3** (A) was monitored by HPLC over time at 25 °C and pH 7.5. The composition of the different peaks was confirmed by mass spectrometry.

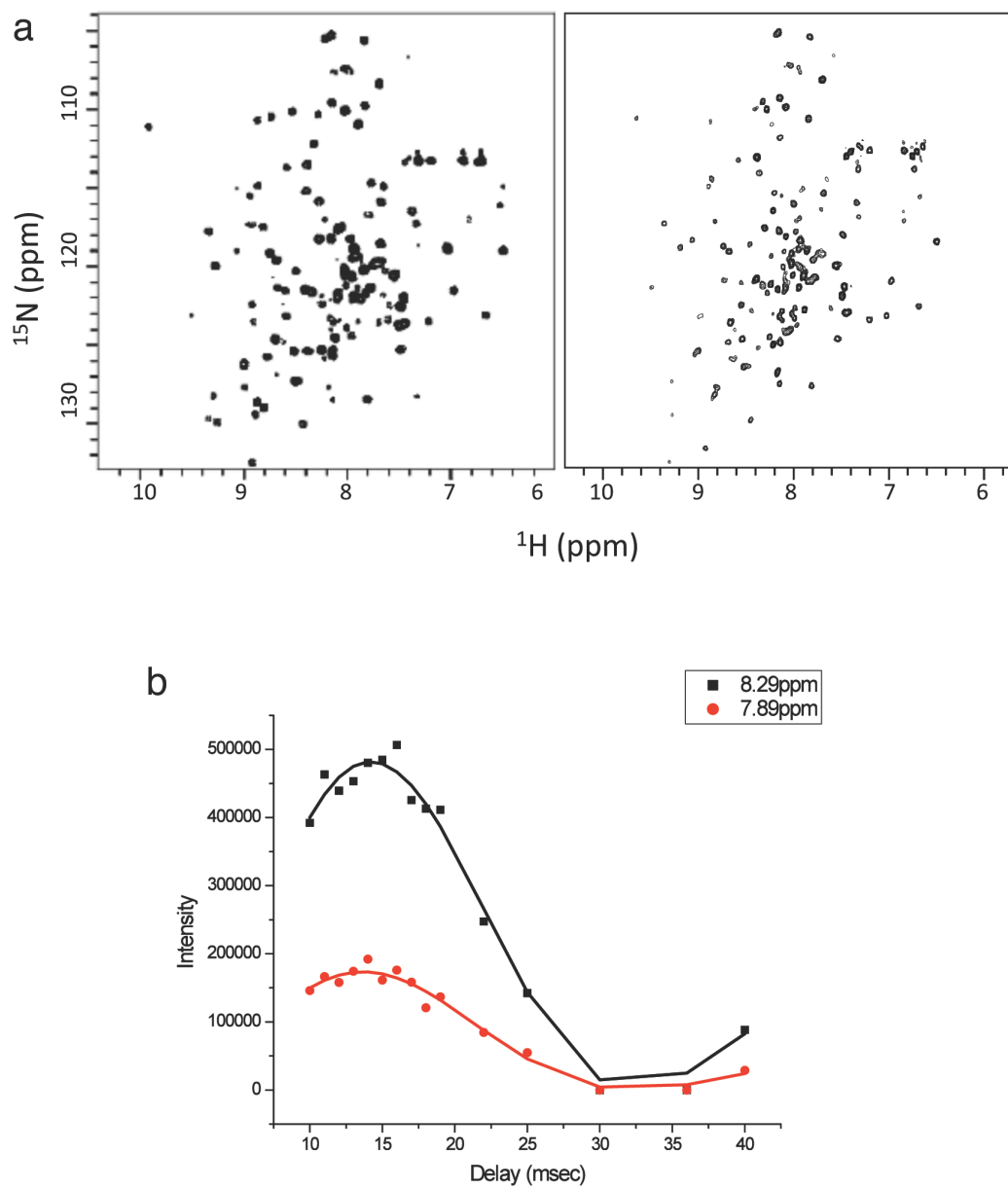


**Fig. S10:**  $^1\text{H}\{^{15}\text{N}\}$  HSQC spectra of refolded constructs **5** (a), **6** (b) and **7** (c). (d) Overlap  $^1\text{H}\{^{15}\text{N}\}$  HSQC spectra of construct **5** (red), **6** (green) and **7** (black).

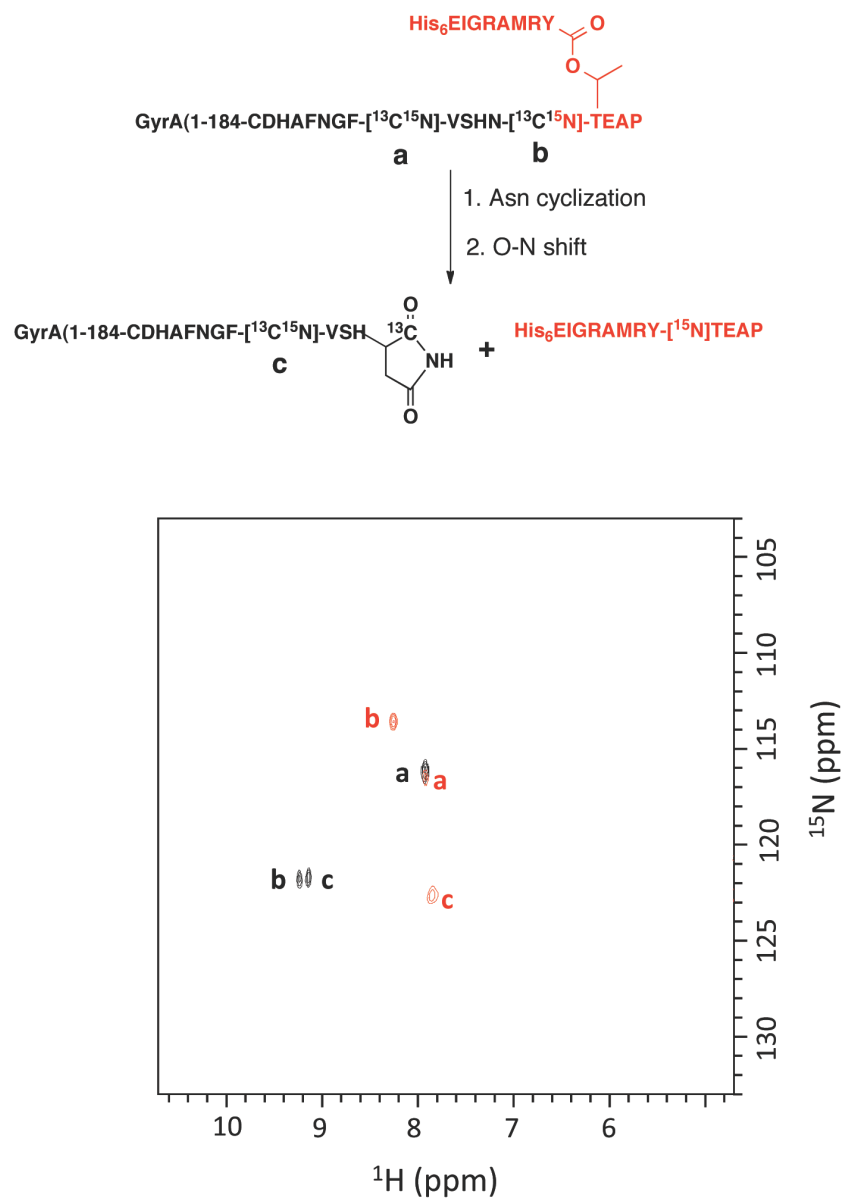




**Fig. S11:** Activity of construct 7. **(a)** Construct 7 was incubated in splicing buffer and aliquots of the mixture periodically analyzed by RP-HPLC. The product of ester hydrolysis accumulated over time. **(b)** ESI-MS analysis of ester hydrolysis product.



**Fig. S12:** (a)  $^1\text{H}\{^{15}\text{N}\}$  HSQC spectra of refolded constructs **10** (right), and **11** (left). (b) Determination of the  $^1J_{\text{NC}}$  coupling constant of the amide bond +1 (black) and amide bond Phe 194-Val195 (red).



**Fig. S13:** Solution  $^1\text{H}\{^{15}\text{N}\}$  HNCO spectrum of construct **12** after incubation at pH 7.5 for ~8 hours (black). Under these conditions we observe signals from both the starting material (labeled a, b) and the succinimide product (labeled c). Subsequently, the sample was acidified to pH 4.5 and the  $^1\text{H}\{^{15}\text{N}\}$  HNCO spectrum re-acquired (red). Both spectra were recorded on a 900 MHz spectrometer at 25°C.